

- 1 1. Method for automatic community model generation based on uni-parity data,
2 comprising the steps of:
3 hypothesizing a subset S of set U, wherein for any pair of items in said
4 subset S there exists a mathematical function C applicable to said pair of items so
5 as to generate a correlation value and correlation relationship between any said
6 pair of items in subset S;
7 generating said correlation values by applying said function C to each of
8 said pairs of items in said subset S;
9 graphing G(S,E), wherein E is the edge set of said graph G with computed
10 correlation values as weights; and
11 mapping said graph G to one of its subgraphs $M \subseteq G$ so as to generate a
12 community.
13
- 1 2. Method of claim 1, wherein said correlation relationship and said correlation
2 value is defined by:
3 $\forall p, q \in S \subseteq U, C : S \times S \rightarrow [0,1]$
4 and wherein said correlation value is in the range of [0,1].
5
- 1 3. Method for solving a community generation problem, comprising the steps of:
2 converting documents to digital form and tagging said digitized
3 documents;
4 parsing said digitized and tagged documents to extract the transaction
5 history vector for each individual;
6 creating timelines of said transaction vectors so as to form a timeline map;
7 determining the relevancy of said vectors;
8 projecting said vectors along a time dimension so as to form a histogram;
9 translating said vectors into groups of activities by histogram clustering;
10 determining the local correlation between any pair of clusters in the
11 timeline of two individuals;
12 computing the global correlations between pairs of individuals;

13 converting data to a graph as a function of all individuals extracted from
14 said documents and the correlation values between said individuals;
15 generating models based on a search of all subgraphs with correlation
16 values above a threshold; and
17 outputting a group model.

18
1 4. Method of claim 3, wherein said step of parsing further comprises the step of
2 applying the “one way nearest neighbor” principle.

3
1 5. Method of claim 4, wherein said “one way nearest neighbor” principle further
2 comprises the following steps as applied to a money laundering problem:

3 for every person’s name encountered, the first immediate time instance
4 is the first time instance for a series of financial activities; the second
5 immediate time instance is the second time instance for another series of
6 financial activities, etc.;

7 for every time instance encountered, all the subsequent financial
8 activities are considered as the series of financial activities between this time
9 instance and the next time instance;

10 financial activities are identified in terms of money amount; money
11 amount is neutral in terms of deposit or withdrawal;

12 each person’s time sequence of financial activities is updated if new
13 financial activities of this person are encountered in other places of the same
14 document or in other documents; and

15 the financial activities of each time instance of a person is updated if
16 new financial activities of this time instance of the same person are
17 encountered in other places of the same document or in other documents.

18
1 6. Method of claim 3, wherein said step of determining the relevancy of said
2 vectors further comprises a step of focusing on “clusters” of vectors in said
3 timeline map and ignoring scattered (i.e., non-clustered) vectors in said timeline
4 map.

7. Method of claim 3, wherein said step of translating said vectors into groups of activities further comprises solving a standard histogram clustering problem; and simplifying said standard clustering problem by virtue of all individuals sharing the same said timeline.

8. Method of claim 3, wherein said step of computing correlations between pairs of individuals further comprises computing the global correlation of all local correlations between pairs of individuals.

9. Method of claim 8, further comprising the step of computing local correlations by computing the correlation between two clusters corresponding to a pair of individuals on said histograms.

10. Method of claim 9, wherein said step of computing correlations between two clusters further comprises the step of computing the fuzzified correlation between $fx_i(t)$ and $fy_j(t)$, the financial transaction histogram functions of individual x and y in cluster i and j , respectively.

11. Method of claim 10, wherein said step of computing the fuzzified correlation between $fx_i(t)$ and $fy_j(t)$ further comprises the step of computing the maximum correlation value

$$g(x_i, y_j) = \max_{i=0}^{w_i} \sum_{t'=0}^{w_j} g x_i(t') g y_j(t - t')$$

where $g x_i(t) = \sum_{t'=1}^{w_i} f x_i(t') G_i(t, t')$

where $G_i(a, b) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(a-b)^2}{2\sigma_i^2}}$

and where $\sigma_i = \frac{2}{W_i(W_i - 1)} \sum_{a=1}^{w_i} \sum_{b=a+1}^{w_i} |a - b|$

12. Method of claim 8 wherein said step of computing the global correlation of all local correlations between pairs of individuals further comprises computing the dot product between two vectors as follows:

$$C(x, y) = Cy(x) \bullet Cx(y) = \sum_{i=1}^K C(x_i, y)C(y_i, x)$$

where the vectors $Cy(x)$ and $Cx(y)$ are defined as

$$Cy(x) = \langle C(x_i, y), i = 1, \dots, K \rangle$$

$$Cx(y) = \langle C(y_i, x), i = 1, \dots, K \rangle$$

where $C(x_i, y) = \max_{j=1}^K \{g(x_i, y_j)S(i, j)\}$

$$S(i, j) = e^{-\frac{(c_i - c_j)^2}{2\sigma_i^2}}$$

and where

$$\{g(x_i, y_j)S(i, j), j = 1, \dots, K\}$$

13. Method of claim 3, wherein said step of converting data to a graph further comprises obtaining a complete graph $G(V, E)$, where V is the set of all the individuals extracted from the given collection of the documents, and E is the set of all the correlation values between individuals such that for any correlation $C(x, y)$, there is a corresponding edge in G with the weight C between the two nodes x and y .

14. Method of claim 3, wherein said step of generating models further comprises the step of identifying links as a graph segmentation based on a minimum correlation threshold value.

1 15. Method of claim 14, wherein said minimum threshold value is selected based
2 upon a user's expertise.